

**NATIONAL BUREAU OF STANDARDS REPORT****NBS PROJECT**

1103-40-11625

9 June 1961

**NBS REPORT**

7167

**ON THE "SYNTHETIC RECORD" PROBLEM**

(Autoregressive Models for Stream Discharge)

by

Joan Raup Rosenblatt

Statistical Engineering Laboratory

**Technical Report No. 3**

to

Water Resources Division  
U. S. Geological Survey  
Department of Interior**IMPORTANT NOTICE**

NATIONAL BUREAU OF ST/ intended for use within the / to additional evaluation and r/ listing of this Report, either i/ the Office of the Director, Na/ however, by the Government / to reproduce additional copies. . . . .

Approved for public release by the  
director of the National Institute of  
Standards and Technology (NIST)  
on October 9, 2015

Progress accounting documents  
nally published it is subjected  
eproduction, or open-literature  
ion is obtained in writing from  
Such permission is not needed,  
prepared if that agency wishes

**U. S. DEPARTMENT OF COMMERCE****NATIONAL BUREAU OF STANDARDS**



ON THE "SYNTHETIC RECORD" PROBLEM  
(Autoregressive Models for Stream Discharge)

Joan Raup Rosenblatt

1. Introduction

This report is supplementary to NBS Report 6513 (July 1959), in which the background for the problem and a simple probabilistic model are described. The model stated in that report is now modified to incorporate autocorrelations in the time series of stream discharges.

In this study, the correlation structure for a bivariate time series is derived from an autoregressive model (Section 2). A given correlation structure does not, however, uniquely determine a bivariate autoregressive scheme. At the end of Section 2, the results are given of a brief study of possible autoregressive schemes consistent with the correlation structure derived from the model assumed here. This may give some insight into the interpretation of the assumptions and their applicability in hydrologic studies.

The main results of this investigation describe the properties of a procedure for estimating the mean  $\mu_y$  of the distribution of discharge for a stream for which there is a "short record", the estimation procedure making use of the correlation between the discharge from this stream and the discharge from a nearby stream for which there is a "long record". Estimation of the mean discharge  $\mu_y$  is discussed in Section 3. Derivation of the variance formula is outlined in an appendix.



## 2. Autoregressive Models for Stream Discharge

### 2.1 Derivation of Correlation Structure

It is assumed, as in the earlier report, that the simultaneous discharges  $(X_t, Y_t)$  from two streams at time  $t$  have a joint normal distribution with parameters  $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho = \rho_{xy} = \rho_x \rho_y$ . Stream discharge is assumed to be stationary; i.e., these parameters (among others) are the same for any  $t = 0, \pm 1, \pm 2, \dots$ .

Instead of assuming independence among pairs  $(X_t, Y_t)$  at different times, it is assumed that each of the two discharge series  $\{X_t\}$  and  $\{Y_t\}$  is dependent, with autocorrelations arising from an autoregressive structure of the form:

$$X_t - \mu_x = \rho_x (X_{t-1} - \mu_x) + U_t,$$

$$Y_t - \mu_y = \rho_y (Y_{t-1} - \mu_y) + V_t,$$

where  $\{(U_t, V_t)\}$  is a sequence of bivariate random variables which are assumed to be independent and identically distributed with joint normal distribution. Also, any  $U_t$  or  $V_t$  is independent of any  $X_{t-s}$  or  $Y_{t-s}$  ( $s > 0$ ).

From the assumptions about the joint distribution of  $(X_t, Y_t)$ , it follows that  $U_t$  and  $V_t$  have zero means, variances

$$\sigma_u^2 = (1 - \rho_x^2) \sigma_x^2,$$

$$\sigma_v^2 = (1 - \rho_y^2) \sigma_y^2,$$

and correlation

$$\rho_{uv} = \rho(1 - \rho_x \rho_y) / \sqrt{1 - \rho_x^2} \sqrt{1 - \rho_y^2}.$$



In order that the above formulas make sense, it is necessary that

$$\rho_x^2 \leq 1, \quad \rho_y^2 \leq 1, \quad \text{and} \quad \rho_{uv}^2 \leq 1, \quad \text{i.e.,}$$

$$\rho^2(1-\rho_x\rho_y)^2 \leq (1-\rho_x^2)(1-\rho_y^2).$$

The last inequality is evidently satisfied if  $\rho_x = \rho_y$ , but places some restrictions on possible values of  $\rho_x$ ,  $\rho_y$ , and  $\rho$  when  $\rho_x \neq \rho_y$ .

We now calculate the lagged correlations:

$$\text{Corr. } (X_t, X_{t-s}) = \rho_x^s, \quad s > 0,$$

$$\text{Corr. } (Y_t, Y_{t-s}) = \rho_y^s, \quad s > 0,$$

$$\text{Corr. } (X_t, Y_{t-s}) = \rho \rho_x^s, \quad s > 0,$$

$$\text{Corr. } (X_t, Y_{t+s}) = \rho \rho_y^s, \quad s > 0.$$

This completes the derivation of the correlation structure for a sequence of pairs  $\{(X_t, Y_t)\}$ .

One further specializing assumption will be made in the derivations given in Section 3. That is, it will be assumed that

$$\rho_x = \rho_y = \tau.$$

This greatly simplifies the calculations.

## 2.2. Alternative Autoregressive Models

It was remarked in the introduction that the autoregressive structure assumed above is not the only one which might lead to this set of correlations. For a brief general discussion of this question, see for example Quenouille (1957).



As an illustration, the bivariate autoregressive series generated by the following equations would have the same correlation structure:

$$\begin{aligned}
 X_t - \mu_x &= \rho_x (X_{t-1} - \mu_x) \\
 &+ \sigma_x \sqrt{1-\rho_x^2} \sqrt{\frac{1+\rho_{uv}}{2}} \epsilon_t \\
 &+ \sigma_x \sqrt{1-\rho_x^2} \sqrt{\frac{1-\rho_{uv}}{2}} \delta_t, \\
 Y_t - \mu_y &= \rho_y (Y_{t-1} - \mu_y) \\
 &+ \sigma_y \sqrt{1-\rho_y^2} \sqrt{\frac{1+\rho_{uv}}{2}} \epsilon_t \\
 &- \sigma_y \sqrt{1-\rho_y^2} \sqrt{\frac{1-\rho_{uv}}{2}} \delta_t,
 \end{aligned}$$

where  $\epsilon_t$ ,  $\delta_t$  are all mutually independent normal random variables with mean zero and variance unity.

The postulated correlation structure would also be generated by the following set of equations (with the same assumptions about  $\epsilon_t$ ,  $\delta_t$ ):

$$\begin{aligned}
 X_t - \mu_x &= \rho_x (X_{t-1} - \mu_x) + \sigma_x \sqrt{1-\rho_x^2} \epsilon_t, \\
 Y_t - \mu_y &= \rho_y (Y_{t-1} - \mu_y) + \rho_{uv} \sigma_y \sqrt{1-\rho_y^2} \epsilon_t \\
 &+ \sqrt{1-\rho_{uv}^2} \sigma_y \sqrt{1-\rho_y^2} \delta_t.
 \end{aligned}$$

The point of these examples is that there would be many ways of writing  $(U_t, V_t)$  as linear functions of  $(\epsilon_t, \delta_t)$ , where  $\epsilon_t$ ,  $\delta_t$  are independent and  $U_t$ ,  $V_t$  have the prescribed variances



and covariance. Unless there are physical reasons for postulating a particular autoregressive structure, the model employed here has certain ambiguities.

### 3. Estimation of the Mean

Suppose that  $X_t$  denotes discharge from the stream with long record, and  $Y_t$  denotes discharge from the stream with short record. The data given, then, are

$$(X_1, Y_1), \dots, (X_{n_1}, Y_{n_1}),$$

and

$$X_{n_1+1}, \dots, X_{n_1+n_2}.$$

Let  $N = n_1 + n_2$  denote the number of time periods covered by the long record.

Estimation of the mean discharge  $\mu_y$  for the stream with a short record is assumed (as in the earlier report) to employ the "synthetic" values

$$\hat{Y}_{n_1+1}, \dots, \hat{Y}_{n_1+n_2},$$

obtained from the estimated regression of  $Y_t$  on  $X_t$ . Thus, the estimation for  $\mu_y$  is

$$U = \bar{Y}_1 + \frac{n_2}{N} b(\bar{X}_2 - \bar{X}_1),$$

where

$$\bar{X}_1 = \sum_{i=1}^{n_1} X_i / n_1,$$

$$\bar{X}_2 = \sum_{i=1}^{n_2} X_{n_1+i} / n_2,$$



$$\bar{Y}_1 = \sum_{i=1}^{n_1} Y_i / n_1 ,$$

and

$$b = \sum_{i=1}^{n_1} (X_i - \bar{X}_1) (Y_i - \bar{Y}_1) / \sum_{i=1}^{n_1} (X_i - \bar{X}_1)^2 .$$

The estimator  $U$  is an unbiased estimator for  $\mu_y$  ;

that is,

$$EU = \mu_y ,$$

where  $E$  denotes mathematical expectation.

An approximate formula for the variance of  $U$  , including terms up to order  $1/n_1^2$  is:

$$\begin{aligned} \text{Var } U &\doteq \frac{\sigma_y^2}{n_1} \left[ \left( \frac{1+\zeta}{1-\zeta} - \frac{2\zeta}{(1-\zeta)^2} - \frac{1}{n_1} \right) \right. \\ &\quad - \frac{n_2}{N} \left\{ \rho^2 \left( \frac{1+\zeta}{1-\zeta} - \frac{2\zeta}{(1-\zeta)^2} \cdot \frac{N+n_1}{N n_1} \right) \right. \\ &\quad \left. \left. - \frac{(1-\rho^2)}{n_1} \cdot \frac{1+\zeta^2}{(1-\zeta)^2} \right\} \right] \end{aligned}$$

This is written in a form which makes it evident that, for  $\zeta = 0$ , this formula agrees with the formula derived by H. A. Thomas (up to the same order of approximation).

#### Acknowledgement

This investigation was made in response to a request by Dr. N. C. Matalas of the USGS Water Resources Division, who has made numerous helpful suggestions during discussions of the work.

#### Reference

Quenouille, M. H. (1957), The Analysis of Multiple Time-Series,  
New York: Hafner.



### Appendix

The basic approach to the derivation of Var U is given, together with the matrix notation employed. The derivation itself is sketched, with lengthy calculations suppressed.

#### A-1. Transformation into a representation in terms of mutually independent random variables.

For  $t = 2, 3, \dots$ , we have the autoregressive relations

$$X_t - \mu_x = \tau(X_{t-1} - \mu_x) + U_t,$$

$$Y_t - \mu_y = \tau(Y_{t-1} - \mu_y) + V_t,$$

where the  $(U_t, V_t)$  are mutually independent bivariate random variables with zero means, Variances

$$\sigma_u^2 = (1 - \tau^2) \sigma_x^2, \quad \sigma_v^2 = (1 - \tau^2) \sigma_y^2,$$

and correlation  $\rho_{uv} = \rho$ .

Now let

$$\xi_1 = (X_1 - \mu_x)/\sigma_x, \quad \eta_1 = (Y_1 - \mu_y)/\sigma_y,$$

and, for  $t \geq 2$ ,

$$\xi_t = U_t / \sigma_x \sqrt{1 - \tau^2}, \quad \eta_t = V_t / \sigma_y \sqrt{1 - \tau^2}.$$

By repeated use of the autoregressive equations, we derive the transformation

$$\left\{ \begin{array}{l} X_1 - \mu_x = \sigma_x \xi_1, \\ X_t - \mu_x = \tau^{t-1} \sigma_x \xi_1 + \sqrt{1 - \tau^2} \sum_{j=2}^t \tau^{t-j} \sigma_x \xi_j, \end{array} \right. t \geq 2,$$

$$\left\{ \begin{array}{l} Y_1 - \mu_y = \sigma_y \eta_1, \\ Y_t - \mu_y = \tau^{t-1} \sigma_y \eta_1 + \sqrt{1 - \tau^2} \sum_{j=2}^t \tau^{t-j} \sigma_y \eta_j, \end{array} \right. t \geq 2,$$



through which the  $X_t$  and  $Y_t$  are expressed in terms of independent bivariate random variables  $(\xi_t, \eta_t)$  having joint normal distribution with zero means, unit variances, and correlation  $\rho$ .

We now make the further transformation

$$\eta_t = \rho \xi_t + \sqrt{1 - \rho^2} \theta_t, \quad t = 1, 2, \dots, n_1,$$

where the  $\theta_t$  are normal with zero means and unit variances, and the  $\theta_t$  and  $\xi_t$  are all mutually independent.

#### A-2. Matrix notation and some preliminary remarks.

The derivations are most efficiently conducted in matrix notation. This section assembles all the definitions of notation.

$I_m$  - identity matrix of order  $m$

$e_m$  - column vector of  $m$  ones

$J_m$  - matrix of order  $m$  all of whose elements are equal to one

$J_{mn}$  -  $(m \times n)$  rectangular matrix of ones

$A'$  - transpose of the matrix  $A$

$\text{tr } A$  - trace of the (square) matrix  $A$ , i.e., the sum of the elements on the diagonal.

Extensive use is made of the fact that  $\text{tr}(AB) = \text{tr}(BA)$ , whenever the products  $AB$  and  $BA$  are both possible. In particular, any quadratic form  $x' Ax$  may be written  $\text{tr}(A x x')$ .

We now introduce the matrix notation for the random variables  $X_t$ ,  $Y_t$ , and for the transformations given in the preceding section. Let

$$x_1 = (X_1, \dots, X_{n_1})',$$

$$x_2 = (X_{n_1+1}, \dots, X_{n_1+n_2})',$$

$$y_1 = (Y_1, \dots, Y_{n_1})',$$



each being a column vector. The transformation may be written

$$\begin{bmatrix} x_1 - \mu_x e_{n_1} \\ x_2 - \mu_x e_{n_2} \end{bmatrix} = \sigma_x \begin{bmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix},$$

$$y_1 - \mu_y e_{n_1} = \sigma_y G_{11} \eta_1,$$

where (at the risk of minor confusion),

$$\xi_1 = (\xi_1, \dots, \xi_{n_1})'$$

$$\xi_2 = (\xi_{n_1+1}, \dots, \xi_{n_1+n_2})'$$

$$\eta_1 = (\eta_1, \dots, \eta_{n_1})'.$$

Also,

$$\eta_1 = \rho \xi_1 + \sqrt{1 - \rho^2} \theta_1,$$

with

$$\theta_1 = (\theta_1, \dots, \theta_{n_1})'.$$

Now, define

$$S = \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2.$$

In the matrix notation, this becomes

$$S = x_1' (n_1 I_{n_1} - J_{n_1}) x_1$$

$$= \sigma_x^2 \xi_1' G_{11}' (n_1 I_{n_1} - J_{n_1}) G_{11} \xi_1,$$

and it will be convenient to define

$$T = G_{11}' (n_1 I_{n_1} - J_{n_1}) G_{11},$$

so that

$$S = \sigma_x^2 \xi_1' T \xi_1.$$

Now the estimated regression coefficient may be written (recalling  $\beta = \rho \sigma_y / \sigma_x$ ),



$$\begin{aligned}
 b &= \frac{1}{S} \mathbf{x}_1' (\mathbf{n}_1 \mathbf{I}_{n_1} - \mathbf{J}_{n_1}) \mathbf{Y}_1 = \frac{\sigma_x \sigma_y}{S} \xi_1' \mathbf{T} \eta_1 \\
 &= \frac{\rho \sigma_x \sigma_y}{S} \xi_1' \mathbf{T} \xi_1 + \sqrt{1 - \rho^2} \frac{\sigma_x \sigma_y}{S} \xi_1' \mathbf{T} \theta_1 \\
 &= \beta + \sqrt{1 - \rho^2} \frac{\sigma_x \sigma_y}{S} \xi_1' \mathbf{T} \theta_1.
 \end{aligned}$$

Finally, the estimator for  $\mu_y$  may be written

$$\begin{aligned}
 U &= \bar{Y}_1 + \frac{n_2}{N} b(\bar{X}_2 - \bar{X}_1) \\
 &= \bar{Y}_1 + \frac{n_2}{N} \beta(\bar{X}_2 - \bar{X}_1) + \frac{n_2}{N} (b - \beta)(\bar{X}_2 - \bar{X}_1).
 \end{aligned}$$

To verify that  $EU = \mu_y$ , we observe that  $E \bar{Y}_1 = \mu_y$ , and  $E(\bar{X}_2 - \bar{X}_1) = 0$ . Also, it has been established above that  $(b - \beta)$  is a linear function of the  $\theta_i$ , which are independent of the  $\xi_i$ ; and  $(\bar{X}_2 - \bar{X}_1)$  is a linear function of the  $\xi_i$ . Hence  $E(b - \beta)(\bar{X}_2 - \bar{X}_1) = 0$ .

### A-3. Derivation of the variance of U.

The derivation involves some lengthy calculations, which will not be given in full. Referring to the expression for  $U$  given at the end of the preceding section, we write

$$\begin{aligned}
 U - \mu_y &= (\bar{Y}_1 - \mu_y) + \frac{n_2}{N} \beta (\bar{X}_2 - \bar{X}_1) \\
 &\quad + \frac{n_2}{N} (b - \beta)(\bar{X}_2 - \bar{X}_1),
 \end{aligned}$$

and

$$\begin{aligned}
 (U - \mu_y)^2 &= (\bar{Y}_1 - \mu_y)^2 + \frac{n_2^2}{N^2} \beta^2 (\bar{X}_2 - \bar{X}_1)^2 \\
 &\quad + 2 \frac{n_2}{N} \beta (\bar{X}_2 - \bar{X}_1) (\bar{Y}_1 - \mu_y) \\
 &\quad + 2 \frac{n_2}{N} (b - \beta)(\bar{X}_2 - \bar{X}_1) (\bar{Y}_1 - \mu_y) \\
 &\quad + \frac{n_2^2}{N^2} (b - \beta)^2 (\bar{X}_2 - \bar{X}_1)^2 + \frac{2 n_2^2}{N^2} \beta (b - \beta) (\bar{X}_2 - \bar{X}_1)^2
 \end{aligned}$$



The last of the six terms in the expression above is a linear function of the  $\theta_i$  which enter only in  $(b - \beta)$ , and hence has expected value zero. The expected values of the first three terms are easily obtained, and are

$$E(\bar{Y}_1 - \mu_y)^2 = \frac{\sigma_y^2}{n_1^2} e'_{n_1} G_{11} G_{11}' e_{n_1},$$

$$\frac{n_2^2}{N^2} \beta^2 E(\bar{X}_2 - \bar{X}_1)^2 = \frac{\beta^2 \sigma_x^2}{N^2} \left\{ \frac{n_2^2}{n_1^2} e'_{n_1} G_{11} G_{11}' e_{n_1} + e'_{n_2} G_{21} G_{21}' e_{n_2} + e'_{n_2} G_{22} G_{22}' e_{n_2} - 2 \frac{n_2}{n_1} e'_{n_2} G_{21} G_{11}' e_{n_1} \right\},$$

$$2 \frac{n_2}{N} \beta E(\bar{X}_2 - \bar{X}_1)(\bar{Y}_1 - \mu_y) = \frac{2 \beta^2 \sigma_x^2}{N} \left\{ \frac{1}{n_1} e'_{n_2} G_{21} G_{11}' e_{n_1} - \frac{n_2}{n_1^2} e'_{n_1} G_{11} G_{11}' e_{n_1} \right\}.$$

If we write

$$G = \begin{bmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{bmatrix},$$

then, after some rearranging, the sum of the three terms given above is:

$$\frac{\rho^2 \sigma_y^2}{N^2} e'_{N} G G' e_{N} + (1-\rho^2) \frac{\sigma_y^2}{n_1^2} e'_{n_1} G_{11} G_{11}' e_{n_1}$$

Next, consider the fourth and fifth terms in the expression for  $(U - \mu_y)^2$ . These involve  $S$  and  $S^2$  in their denominators, and their expected values will be evaluated approximately. First, however, observe that the vectors  $\theta_1$  and  $\xi_2$  appear only in the numerators. Holding the vector  $\xi_1$  fixed, and taking



expectations with respect to the random variables in  $\theta_1$  and  $\xi_2$ , the calculations for the fourth and fifth terms are reduced to the following:

$$\begin{aligned} \frac{2n_2}{N} E(b-\beta)(\bar{X}_2 - \bar{X}_1)(\bar{Y}_1 - \mu_y) &= \frac{2}{N} \sigma_y^2 (1 - \rho^2) \frac{n_2}{n_1} \sigma_x^2 \\ E \frac{1}{S} \left[ \frac{1}{n_2} \xi_1' G_{21}' J_{n_2 n_1} G_{11} T \xi_1 - \frac{1}{n_1} \xi_1' G_{11}' J_{n_1} G_{11} T \xi_1 \right] \\ \frac{n_2^2}{N^2} E (b-\beta)^2 (\bar{X}_2 - \bar{X}_1)^2 &= \frac{n_2^2}{N^2} \sigma_y^2 (1 - \rho^2) \sigma_x^4 \\ E \frac{1}{S^2} \xi_1' T^2 \xi_1 \left[ \frac{1}{n_2^2} e' n_2 G_{22} G_{22}' e_{n_2} \right. \\ + \frac{1}{n_1^2} \xi_1' G_{11}' J_{n_1} G_{11} \xi_1 + \frac{1}{n_2^2} \xi_1' G_{21}' J_{n_2} G_{21} \xi_1 \\ \left. - \frac{2}{n_1 n_2} \xi_1' G_{11} J_{n_1 n_2} G_{21} \xi_1 \right] \end{aligned}$$

There are six terms to be evaluated here, each involving a ratio of random variables. For a first-order approximation, each random variable is replaced by its expected value. Further approximation involves deletion of certain terms having the factor  $\tau^{n_1}$  (or  $\tau^{n_2}$ ), which is assumed to be negligible. Finally, only terms of order up to  $1/n_1^2$  are retained.

Evaluation of the first term will be sketched, and the values of the other five terms will be listed. For the first term, we require

$$E \frac{1}{S} \xi_1' G_{21}' J_{n_2 n_1} G_{11} T \xi_1 .$$

Now

$$\begin{aligned} \frac{1}{\sigma_x^2} E S &= E \xi_1' T \xi_1 = E \text{tr}(T \xi_1 \xi_1') \\ &= \text{tr}[T \cdot E(\xi_1 \xi_1')] = \text{tr } T , \end{aligned}$$

and

$$\begin{aligned} \text{tr } T &= \text{tr}[G_{11}' (n_1 I_{n_1} - J_{n_1}) G_{11}] \\ &= n_1 \text{tr } G_{11}' G_{11} - \text{tr } G_{11}' J_{n_1} G_{11} ; \end{aligned}$$



but we can write  $J_{n_1} = e_{n_1} e'_{n_1}$ , whence

$$\text{tr } T = n_1 \text{ tr } G_{11} G_{11}' - e'_{n_1} G_{11} G_{11}' e_{n_1}.$$

Now it is easily verified that

$$G_{11} G_{11}' = \begin{bmatrix} 1 & \tau & \tau^2 & \dots & \tau^{n_1-1} \\ \tau & 1 & \tau & \dots & \tau^{n_1-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tau^{n_1-1} & \tau^{n_1-2} & \tau^{n_1-3} & \dots & 1 \end{bmatrix}.$$

Thus the trace of this matrix is  $n_1$  and the sum of all the elements of this matrix is

$$n_1 \frac{1+\tau}{1-\tau} - \frac{2\tau}{(1-\tau)^2} (1 - \tau^{n_1}).$$

Finally, then, deleting the negligible term in  $\tau^{n_1}$ ,

$$E S \doteq \left[ n_1^2 - n_1 \frac{1+\tau}{1-\tau} + \frac{2\tau}{(1-\tau)^2} \right] \sigma_x^2$$

Next, in a similar way, we obtain

$$\begin{aligned} E \xi' G_{21}' J_{n_2 n_1} G_{11} T \xi_1 &= \text{tr}(G_{21}' J_{n_2 n_1} G_{11} T) \\ &= e'_{n_1} G_{11} T G_{21}' e_{n_2} \\ &= n_1 e'_{n_1} G_{11} G_{11}' G_{11} G_{21}' e_{n_2} \\ &\quad - (e'_{n_1} G_{11} G_{11}' e_{n_1})(e'_{n_1} G_{11} G_{21}' e_{n_2}) \\ &\doteq \frac{2\tau^2}{(1-\tau)^4} - \frac{n_1 \tau^2}{(1-\tau)^3 (1+\tau)}, \end{aligned}$$

after deletion of terms involving  $\tau^{n_1}$  and  $\tau^{n_2}$ .

The ratio of these two expected values is of order  $1/n_1$ , and this ratio enters a term which is already of order  $1/n_1^2$  (actually  $1/n_1 n_2$ ); hence this term will not appear in the



first-order approximation.

The second term and also the last two terms turn out similarly to make no contribution to an approximation up to order  $1/n_1^2$ .

For the third term, we require terms up to order  $1/n_1^2$  in:

$$\begin{aligned}
 & \frac{1}{n_2^2} e_{n_2} G_{22} G_{22}' e_{n_2} \cdot E \frac{1}{S^2} \xi_1' T^2 \xi_1 \\
 & \stackrel{\approx}{=} \frac{1}{n_2^2} \left[ n_2 \frac{1+\tau}{1-\tau} - \frac{2\tau}{(1-\tau)^2} \right] \\
 & \left[ \frac{1+\tau^2}{1-\tau^2} n_1^3 + O(n_1^2) \right] \diagup \sigma_x^4 \left[ n_1^4 + O(n_1^3) \right] \\
 & \stackrel{\approx}{=} \frac{1}{\sigma_x^4} \frac{1+\tau^2}{(1-\tau)^2} \frac{1}{n_1 n_2} .
 \end{aligned}$$

For the fourth term, we have similarly:

$$\begin{aligned}
 & \frac{1}{n_1^2} E \frac{1}{S^2} \xi_1' T^2 \xi_1 \cdot \xi_1' G_{11}' J_{n_1} G_{11} \xi_1 \\
 & \stackrel{\approx}{=} \frac{1}{\sigma_x^4} \frac{1+\tau^2}{(1-\tau)^2} \frac{1}{n_1^2} .
 \end{aligned}$$

Assembling all these terms together with the terms evaluated previously without approximations, we obtain the variance formula given in Section 3 of this report.

